Color Preserving HDR Fusion for Dynamic Scenes

Gökdeniz Karadağ
Middle East Technical University, Turkey
gokdeniz@ceng.metu.edu.tr

Ahmet Oğuz Akyüz
Middle East Technical University, Turkey
akyuz@ceng.metu.edu.tr

ABSTRACT

We present a novel algorithm to efficiently generate high quality high dynamic range (HDR) images. Our method is based on the idea of expanding the dynamic range of a reference image at granularity of tiles. In each tile, we use data from a single exposure, but different tiles can come from different exposures. We show that this approach is not only efficient and robust against camera and object movement, but also improves the color quality of the resulting HDR images. We compare our method against the commonly used HDR generation algorithms.

Keywords: High dynamic range imaging, image fusion, color quality

1 INTRODUCTION

The interest in HDR imaging has rapidly gained popularity in recent years. This has been accompanied by the development of various methods to create HDR images. While it is believed that using dedicated HDR capture hardware will be the de-facto way of generating HDR images in future [Rei10a], software solutions are still commonly used in today's systems. Among these multiple exposure techniques (MET) are the most dominant [Man95a, Deb97a].

In METs, several images of the same scene are captured by varying the exposure time between the images. This ensures that each part of the captured scene is properly exposed in at least one image. The individual images are then merged to obtain the HDR result. Although variations exist, the equation below is typically used for the merging process:

$$I_j = \sum_{i=1}^{N} \frac{f^{-1}(p_{ij})w(p_{ij})}{t_i} \Big/ \sum_{i=1}^{N} w(p_{ij}). \qquad (1)$$

Here N is the number of LDR images, p_{ij} is the value of pixel j in image i, f is the camera response function, t_i is the exposure time of image i, and w is a weighting function used to attenuate the contribution of poorly exposed pixels.

In Equation 1, a weighted average is computed for every pixel. While this may be desirable for attenuating noise, it introduces unwanted artifacts due to ghosting and misalignment problems. In this paper, we show that this approach also results in the desaturation of colors making the HDR image less saturated than the its constituent exposures.

Computing a weighted average for every pixel also requires that the individual pixels are perfectly aligned. Otherwise, pixels belonging to different regions in the scene will be accumulated resulting ghosting and alignment artifacts.

In this paper, we propose a method that largely avoids both of these problems. Our method is underpinned by the idea that instead of computing an average for every pixel, one can use the pixels from a single properly exposed image. A different image can be used for different regions ensuring that the full dynamic range is captured. We also introduce the concept of working in tiles instead of pixels to make the algorithm more robust against local object movements.

2 PREVIOUS WORK

Starting with the pioneering works of Madden [Mad93a] and Mann and Picard [Man95a], various algorithms have been developed to create HDR images. The early work focused on recovering the camera response function and choosing an appropriate weighting function [Deb97a, Mit99a, Rob03a, Gro04a]. These algorithms assumed that the exposures that are used to create an HDR image are perfectly aligned and the scene is static.

Ward developed a method based on median threshold bitmaps (MTBs) to allow photographers use hand-held images of static scenes in HDR image generation [War03a]. His alignment algorithm proved to be very successful and is used as an initial step of more advanced alignment and ghost removal algorithms [Gro06a, Jac08a, Lu09a].

In another alignment algorithm, Cerman and Hlaváč estimated the initial shift amounts by computing the correlation of the images in the Fourier domain [Cer06a]. This, together with the initial rotational estimate which

was assumed to be zero, was used as a starting point for the subsequent iterative search process.

Tomaszewska and Mantiuk employed a modified scale invariant feature transform (SIFT) [Low04a] to extract local features in the images to be aligned [Tom07a]. The prominent features are then selected by the RANSAC algorithm [Fis81a]. This refined set of features are then used to compute a homography between the input images.

Several methods have been proposed to deal with ghosting artifacts. These algorithms usually pre-align the input exposures using MTB or other algorithms to simplify the ghost detection process. Some of these algorithms avoid merging suspicious regions where there is high variance [Kha06a, Gal09a, Ram11a]. Other algorithms try to detect the movement of pixels and perform pixel-wise alignment [Zim11a]. A recent review of HDR ghost removal algorithms can be found in Srikantha and Sidibé [Sri12a].

There are also existing algorithms that attempt to combine data from multiple exposures for the purpose of generating a single low dynamic range (LDR) image. Among these, Goshtasby first partitions the images into tiles [Gos05a]. For each tile, he then selects the image that has the highest entropy. The tiles are blended using smooth blending functions to prevent seams. Mertens et al., on the other hand, do not use tiles but utilize three metrics namely contrast, saturation, and well-exposedness to choose the best image for each pixel [Mer07a]. Similar to Goshtasby, Várkonyi-Kóczy et al. propose a tile based algorithm where tiles are selected to maximize detail using image gradients [Var08a]. In another tile based algorithm, Vavilin and Jo use three metrics; mean intensity, intensity deviation, and entropy to choose the best exposure for each tile [Vav08a]. In contrast to previous tile based studies, they choose tile size adaptively based on local contrast. Finally, Jo and Vavilin propose a segmentation based algorithm which allows choosing different exposures for different clusters [Jo11a]. Unlike previous methods they use bilateral filtering during the blending stage.

It is important to note that existing tile-based algorithms attempt to generate LDR images with more details and enhanced texture information, whereas our goal is to generate HDR images with natural colors. Our approach alleviates the need for explicit ghost detection and removal procedures. If the dynamic parts of a scene do not span across regions with significantly different luminance levels, no ghost effects will occur in the output. Also, we avoid redundant blending of pixels that can result in reduced color saturation.

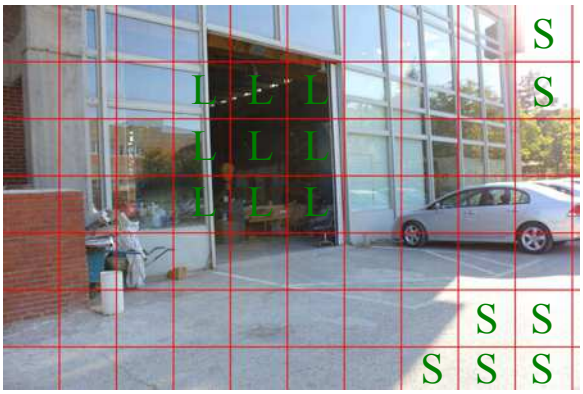


Figure 1: We partition the images into tiles and determine which exposure to use for each tile.

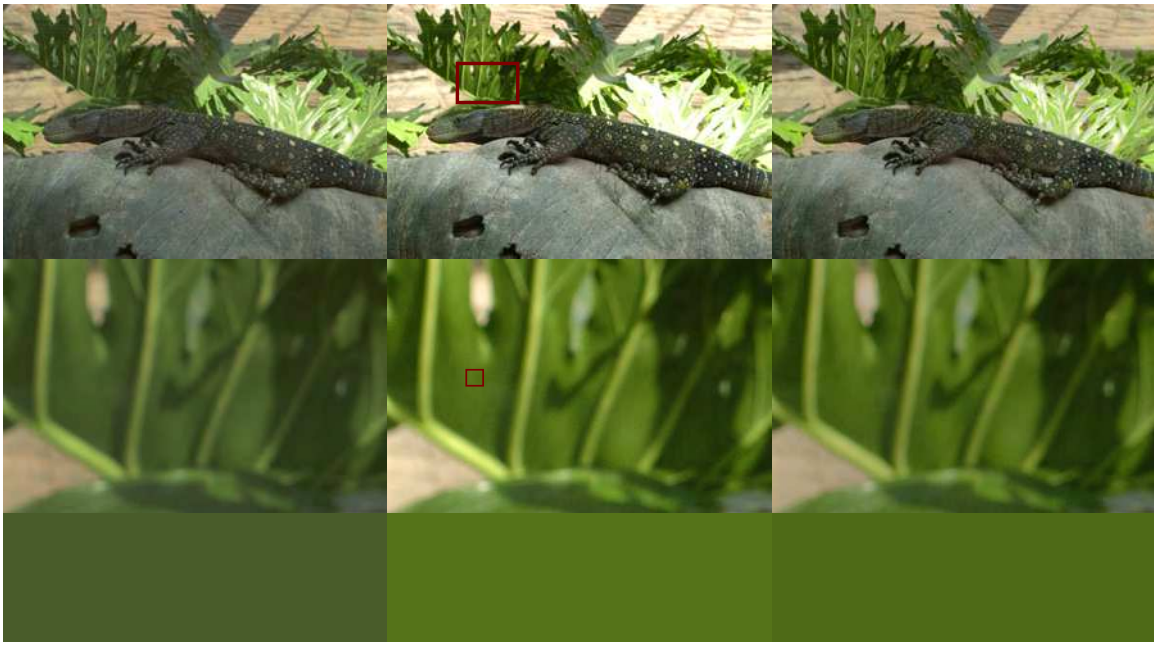
3 ALGORITHM

The first step of our algorithm is to align the input exposures using the MTB algorithm [War03a]. In this part, both the original MTB or the MTB with the rotation support can be used.

Once the images are aligned, we partition each exposure into tiles. Our goal then becomes to choose the best image that represents the area covered by each tile. A sample image is shown in Figure 1 to illustrate this idea. In this image, the under-exposed tiles are marked with **L** indicating that these tiles should come from a longer exposure. Similarly, over-exposed regions are marked by **S** suggesting that shorter exposures should be used for these tiles. Unmarked tiles can come from the middle exposures.

To make these decisions, we need to define a quality metric that indicates whether a tile is well-exposed. To this end, we experimented with the mean intensity as well as the number of under- and over-exposed pixels within a tile as potential metrics. Our results suggested that using the mean intensity gives better results. Therefore, we marked a tile as a *good* tile if its mean intensity is in the range $[I_{min}, I_{max}]$. I_{min} and I_{max} are user parameters, but we found that $I_{min} = 50$ and $I_{max} = 200$ can be used as reasonable defaults.

Based on this criteria, we compute the number of good tiles for each exposure. We choose the exposure with the maximum number of good tiles as the reference exposure. This exposure serves as the *donor* which provides data for all tiles whose mean intensity stays in the aforementioned limits. This leniency allows us to use the same image as much as possible and provides greater spatial coherency. For the remaining tiles, we choose the second reference exposure and fill in the tiles which are valid in this exposure. This process is



(a) Standard MET (b) LDR reference (c) Our result

Figure 2: (a) HDR image created by using the standard MET. (b) Selected individual exposure from the bracketed sequence. (c) HDR image created using our algorithm. The top row shows the full images. The middle row shows the close-up view of a selected region. The bottom row shows the color of a single pixel from the region indicated in the middle row. Both HDR images are tone mapped using the photographic tone mapping operator [Rei02a]. As can be seen in the zoomed views, the color quality of our result is closer to the selected reference image.

recursively executed until a source image is found for all tiles[1]. This process can be represented as:

$$I_j = \sum_{i=1}^{N} \frac{f^{-1}(p_{ij})W_{ij}}{t_i}, \tag{2}$$

$$W_{ij} = \begin{cases} 1 & \text{if pixel } j \text{ comes from image } i, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Note that we no longer have the $w(p_{ij})$ term from Equation 1 as we do not compute a weighted average.

Finally, we use a blending strategy to prevent the visibility of seams at tile boundaries. For this purpose, we create Gaussian pyramids of weights (W_{ij}) and Laplacian pyramids of source images. We then merge the images by using Equation 2 at each level of the pyramid and collapse the pyramid to obtain the final HDR image. We refer the reader to Burt and Adelson's original paper for the details of this process [Bur83a].

Since the tiles are not overlapping our algorithm ensures that within each tile data from only a single source image is used. As we demonstrate in the next section, this improves the color saturation of the resulting HDR images. A second observation is that each tile is spatially coherent. This means that motion related artifacts will not occur within tiles. However, such artifacts can still occur across tiles. Thus our algorithm reduces the effect of motion artifacts but does not completely eliminate them.

[1] It is possible that the a tile is under- or over-exposed in all input images. In this case, we choose the longest exposure if the tile is under-exposed and shortest exposure otherwise.

4 RESULTS AND ANALYSIS

We present the results of our color preserving HDR fusion algorithm under three categories namely: (1) Fixed camera & static scene, (2) hand-held camera & static scene, and (3) hand-held camera & dynamic scene. For the first configuration, we illustrate that the color quality of the HDR image created by our method is superior to the output of the standard HDR fusion algorithm shown in Equation 1. A sample result for this case is depicted in Figure 2 where the output of the standard MET is shown on the left and our result is shown on the right. A selected exposure from the bracketed sequence is shown in the middle for reference.

For the image on the left, we used the tent weighting function proposed by Debevec and Malik [Deb97a]. We used the sRGB camera response function for both images, and a tile size of 64×64 for our result. It can be seen that, due to the pixel-wise averaging process, the output of the standard MET has a washed-out appearance. Our result, on the other hand, is colorimetrically closer to the selected exposure. This is a consequence of avoiding unnecessary blending between images.

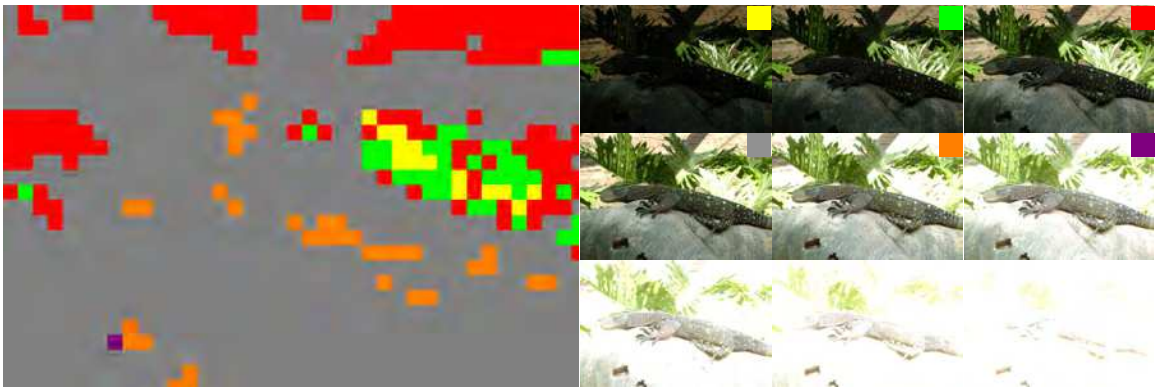


Figure 3: The colors show the correspondence between the tiles in the HDR image and the source images that they were selected from. We can see that most tiles were selected from the fourth image. Figure courtesy of Erik Reinhard [Rei10a].

Figure 3 shows which tiles in the output HDR image came from which images in the exposure sequence. The correspondence is shown by color coding the individual exposures. As we can see from this figure, the majority of the tiles were selected from the fourth exposure. The tiles that correspond to the highlights on the plants came from the darker exposures. On the other hand, the tiles that correspond to the crevices on the rock and the shadow of the lizard came from the lighter exposures. We can also see that the last three exposures were not used at all.

At this point, it would be worthwhile to discuss why the standard MET gives rise to a washed-out appearance and our algorithm does not. We would not expect to see this problem if all exposures were perfect representations of the actual scene. However, in reality, there are slight differences between exposures that are not only due to changing the exposure time. Slight camera movements, noise, and inaccuracies in the camera response curve can all cause variations between the actual observations. The combined effect of these variations result in reduced color saturation. By avoiding unnecessary blending, we also avoid this artifact.

The second test group consists of images of a static scene captured by a hand-held camera (Figure 4). In this figure, the left column shows the unaligned result created by directly merging five bracketed exposures. The middle column shows the tone mapped HDR output after the exposures are aligned by using the MTB algorithm. The right column shows our result obtained by first aligning the exposures using the MTB algorithm, and then merging them using our tile-based technique. As can be seen from the fence and the sign in the insets, our result is significantly sharper than that of the MTB algorithm. However, we also note that small artifacts are visible in our result on the letters "R" and "E". Further examination reveals that these artifacts are due to using tiles from different exposures that are not perfectly aligned.

As the color map indicates, the majority of the final HDR image is retrieved from the exposure coded by red (exposures not shown). The darker regions retrieved data from the lighter (gray) exposure. The highlights at the top left corner received data from the darker (green) exposure. In fact, in this example, all five exposures contributed to the final image but the majority of the contribution came from these three exposures.

In the final category, we demonstrate the performance of our algorithm using scenes that have both global and local movement. To this end, we used the hdrgen software[2] which implements the MTB alignment algorithm and a variance based ghost removal method explained in Reinhard et al. [Rei10a]. In Figure 5, the left column shows the output obtained by only image alignment but without ghost removal. The middle column shows the result of alignment and ghost removal. Although the majority of the ghosts are removed, some artifacts are still visible on the flag as shown in the close-ups. The right column shows our result where these artifacts are eliminated. The color map indicates the source images for different regions of the HDR image.

We also demonstrate a case where our algorithm introduces some unwanted artifacts in high contrast and high frequency image regions as the window example in Figure 6. The bright back light and window grates cause high contrast. If the tile size is large, blending tiles from different exposures produces sub-par results. A reduced tile size eliminates these artifacts.

Our choice of prioritizing the reference image increases success in image sets where ghosting effects would normally occur. If the object movements are located in regions with similar lighting conditions, our algorithm prefers the image closer to reference image while constructing tiles, preventing ghosting effects. It is possible that an object moves between regions of different lighting conditions, and our algorithm may choose tiles

[2] `http://www.anyhere.com`

Figure 4: Left: Unaligned HDR image created from hand-held exposures. Middle: Exposures aligned using the MTB algorithm. Right: Our result. The close-ups demonstrate that our algorithm produces sharper images. The color map shows the source exposures for different regions of the HDR image.

(a) Alignment (b) Alignment and ghost removal (c) Our result

Figure 5: Left: Aligned HDR image created from hand-held exposures using the MTB algorithm. Middle: Aligned and ghost removed HDR image. Right: Our result. The insets demonstrate that ghosting artifacts are eliminated in our result. The color map shows the source exposures for different regions of the HDR image.

from different images where the moving object can be seen. In this case different copies of the object may be present in multiple locations in the output image.

Finally, we report the running times of our algorithm. An unoptimized C++ implementation of our algorithm was able to create high resolution (18 MPs) HDR images from 9 exposures within 30 seconds including all disk read and write times. We conducted all of our test on an Intel Core i7 CPU running at 3.20 GHz and equipped with 6 GBs of memory. This suggests that our algorithm is practical and can easily be integrated into existing HDRI workflows.

5 CONCLUSIONS

We presented a simple and efficient algorithm that improves the quality of HDR images created by using multiple exposures techniques. By not redundantly averaging pixels in low dynamic regions, our algorithm preserves the color saturation of the original exposures, and reduces the effect of ghosting and alignment artifacts. As future work, we are planning to make the tiling process adaptive instead of using a uniform grid. This would prevent artifacts that can be caused by sudden illumination changes between neighboring tiles coming from different exposures. We are also planning to perform blending using edge-aware Laplacian pyramid [Par11a] to avoid blending across sharp edges. Improved quality of our results can also be validated by a user study.

ACKNOWLEDGMENTS

This work was partially supported by METU BAP-08-11-2011-123.

6 REFERENCES

[Bur83a] P. Burt and E. Adelson. The laplacian pyramid as a compact image code. *Communications, IEEE Transactions on*, 31(4):532 – 540, apr 1983.

Figure 6: Top: A tone-mapped HDR image with 128x128 tile size. Tile boundaries are highly visible in the close-up. Middle: Changing the tile size to 32x32 removes most of the artifacts, but some remain in diagonal lines. Bottom: Using a 2x2 tile size eliminates remaining artifacts.

[Cer06a] Lukáš Cerman and Václav Hlaváč. Exposure time estimation for high dynamic range imaging with hand held camera. In *Computer Vision Winter Workshop, Czech Republic*, 2006.

[Deb97a] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH 97 Conf. Proc.*, pages 369–378, August 1997.

[Fis81a] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[Gal09a] O. Gallo, N. Gelfandz, Wei-Chao Chen, M. Tico, and K. Pulli. Artifact-free high dynamic range imaging. In *Computational Photography (ICCP), 2009 IEEE International Conference on*, pages 1 –7, april 2009.

[Gos05a] A. Ardeshir Goshtasby. Fusion of multi-exposure images. *Image and Vision Computing*, 23(6):611 – 618, 2005.

[Gro04a] M.D. Grossberg and S.K. Nayar. Modeling the space of camera response functions. *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, 26(10):1272 –1282, 2004.

[Gro06a] Thorsten Grosch. Fast and robust high dynamic range image generation with camera and object movement. In *Proc. of Vision Modeling and Visualization*, pages 277–284, 2006.

[Jac08a] Katrien Jacobs, Celine Loscos, and Greg Ward. Automatic high-dynamic range image generation for dynamic scenes. *IEEE CG&A*, 28(2):84–93, 2008.

[Jo11a] K.H. Jo and A. Vavilin. Hdr image generation based on intensity clustering and local feature analysis. *Computers in Human Behavior*, 27(5):1507–1511, 2011.

[Kha06a] Erum Arif Khan, Ahmet Oğuz Akyüz, and Erik Reinhard. Ghost removal in high dynamic range images. *IEEE International Conference on Image Processing*, 2006.

[Low04a] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[Lu09a] Pei-Ying Lu, Tz-Huan Huang, Meng-Sung Wu, Yi-Ting Cheng, and Yung-Yu Chuang. High dynamic range image reconstruction from hand-held cameras. In *CVPR*, pages 509–516, 2009.

[Mad93a] B. C. Madden. Extended dynamic range imaging. Technical report, GRASP Laboratory, Uni. of Pennsylvania, 1993.

[Man95a] S Mann and R Picard. Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures, 1995.

[Mer07a] T. Mertens, J. Kautz, and F. Van Reeth. Exposure fusion. In *Computer Graphics and Applications, 2007. PG '07. 15th Pacific Conference on*, pages 382 –390, 29 2007-nov. 2 2007.

[Mit99a] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *Proceedings of CVPR*, volume 2, pages 374–380, June 1999.

[Par11a] Sylvain Paris, Samuel W. Hasinoff, and Jan Kautz. Local laplacian filters: edge-aware image processing with a laplacian pyramid. *ACM Trans. Graph.*, 30(4):68:1–68:12, July 2011.

[Ram11a] Shanmuganathan Raman and Subhasis Chaudhuri. Reconstruction of high contrast images for dynamic scenes. *The Visual Computer*, 27(12):1099–1114, 2011.

[Rei02a] Erik Reinhard, Michael Stark, Peter Shirley, and Jim Ferwerda. Photographic tone reproduction for digital images. *ACM Transactions on Graphics*, 21(3):267–276, 2002.

[Rei10a] Erik Reinhard, Greg Ward, Sumanta Pattanaik, and Paul Debevec. *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*. Morgan Kaufmann, San Francisco, second edition edition, 2010.

[Rob03a] Mark A. Robertson, Sean Borman, and Robert L. Stevenson. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging 12(2), 219 228 (April 2003).*, 12(2):219–228, 2003.

[Sri12a] Abhilash Srikantha and Désiré Sidibé. Ghost detection and removal for high dynamic range images: Recent advances. *Signal Processing: Image Communication*, (0):–, 2012.

[Tom07a] Anna Tomaszewska and Radoslaw Mantiuk. Image registration for multi-exposure high dynamic range image acquisition. In *WSCG: Proc. of the 15th Intl. Conf. in Central Europe on Computer Graphics, Visualization and Computer Vision*, 2007.

[Var08a] A. R. Varkonyi Koczy, A. Rovid, and T. Hashimoto. Gradient-based synthesized multiple exposure time color hdr image. *Instrumentation and Measurement, IEEE Transactions on*, 57(8):1779 –1785, aug. 2008.

[Vav08a] A. Vavilin and K.H. Jo. Recursive hdr image generation from differently exposed images based on local image properties. In *Control, Automation and Systems, 2008. ICCAS 2008. International Conference on*, pages 2791–2796. IEEE, 2008.

[War03a] Greg Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of Graphics Tools*, 8(2):17–30, 2003.

[Zim11a] Henning Zimmer, Andrés Bruhn, and Joachim Weickert. Freehand hdr imaging of moving scenes with simultaneous resolution enhancement. *Computer Graphics Forum*, 30(2):405–414, 2011.